

1st Joint NMFS/UW/OSU workshop series in stock assessment methods

Ray Hilborn & James Ianelli

January 25th – 29th, 1999

Alaska Fisheries Science Center
7600 Sand Pt Way NE
Seattle WA 98115

Draft Syllabus

Day 1: Likelihoods	
Binomial mark recapture	
Regression, normal	
Likelihood profile on regression	
Likelihoods for catch-at-age data	
Spawner recruit and Biomass dynamics models	
Likelihood profile on derived parameters (current stock size)	
Day 2: Bayesian analysis	
Introduction to decision analysis and decision tables	
Lecture on Bayes Law, Monte Hall, Smiths Children, rumpole, lost wallet, etc.	
Bayesian analysis by direct search	
Bayesian linear regression	
Single parameter model	
2 parameter model	
Importance of priors and non-informative priors	
Day 3: Methods for integration SIR and MCMC	
SIR	
MCMC	
Adding alternative policies	
Participants implement their own model using Sir or MCMC	
Comparison to bootstrap	
Reemphasize difference between Bayesian and likelihood approaches	
Day 4: Further topics and examples	
Formulating priors	
Appropriate likelihoods for age and length data	
Decision tables and methods off presenting output	
Day 5: Application to some sample regional problems	
Group examples: delay difference for sablefish? Or full age structured model with indices of abundance but deterministic recruitment	

Workshop Timetable

Time	Activity
900 – 930	Review materials, outline
930 – 1030	Lecture
1030 – 1045	Break
1045 – 1200	Lecture/hands on
1200 – 1315	Lunch
1315 – 1400	Lecture
1400 – 1500	Hands on
1500 – 1700	Lab work, model development, problem solving

1 Likelihood estimation with a simple model

1.1 Purpose

Understand fundamental concept of likelihood estimation in a simple one-parameter model: a population with constant recruitment and known survival. Evaluate information data provide via likelihood inference about recruitment levels, R . Compute profile likelihood and compare with inference derived from standard quadratic approximations of covariance matrix.

1.2 Themes

- Likelihood estimation
- Information content in data
- Likelihood inference via profile likelihood vs. approximation of var-cov matrix
- Fitting process versus observation error models
- Concentrated likelihood vs long-hand

1.3 Data

Population size (in numbers) and catch (in numbers) for every year.

Year	Observed Catch	Biomass from surveys
1971	0	973
1972	6	1054
1973	6	581
1974	6	775
1975	18	942
1976	38	667
1977	68	681
1978	6	547
1979	19	554
1980	22	636
1981	27	657
1982	26	534
1983	32	671
1984	42	681
1985	36	723
1986	86	666
1987	162	619
1988	243	508
1989	198	461
1990	186	306
1991	179	267
1992	155	238
1993	133	361
1994	133	340
1995	105	457
1996	155	329

1.4 The model

Variables and parameters

N_t^o	Observed population in time t
N_t	Predicted population in time t
C_t	Observed harvest in time t
S	Finite survival rate (annual)
R	Average annual recruitment
σ_w^2	Variance of the process error
σ_v^2	Variance of the observation error
n	Number of observations

1.4.1 Dynamic model

$$N_{t+1} = sN_t + R - C_t \quad (1)$$

1.4.2 Initial Conditions

$$N_0 = \frac{R}{(1-S)} \quad (2)$$

1.4.3 Observation error

(variability due to mismeasurement of true state of nature)

$$N_t^o = N_t \exp(v_t); \quad v \sim N(0, \sigma_v^2) \quad (3)$$

1.4.4 Adding process error

(comprises all the variability in the state of the population produced by factors not included in the model).

$$N_{t+1} = [sN_t + R - C_t] \exp(w_{t+1}); \quad w \sim N(0, \sigma_w^2) \quad (4)$$

1.5 Estimation

1.5.1 Likelihood for observation error

The individual observation errors are

$$v_t = \ln \left(\frac{N_t^o}{N_t} \right)$$

The likelihood of the observation error is

$$P(v_t|R) = \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{v_t^2}{2\sigma_v^2}\right]$$

and the likelihood of all observation errors is therefore the product of the likelihoods for all years.

$$P(v|R) = \prod_t \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{v_t^2}{2\sigma_v^2}\right]$$

Finally we note that we since the v 's are the traditional sum of squares, we can convert the sum of squares to the likelihood as follows

$$SSQ = \sum_t v_t^2$$

$$P(v|R) \approx \exp\left(-\frac{SSQ}{2\sigma_v^2}\right)$$

The key to this simplification is that the first term of the likelihood is constant and therefore does not affect the likelihood. Assuming observation error, we can now refer to the likelihood with reference to the observed data.

$$P(N^0|R) \approx \exp\left(\frac{SSQ}{2\sigma_v^2}\right)$$

1.6 Homework: fitting biomass dynamics

On the attached sheet you will find a the history of catches and an index of abundance (total egg deposition from surveys) for herring in the Gulf of Georgia, British Columbia.. Fit the Fox form of the biomass dynamics model to these data using the observation error model. Assume log normal error, and use the analytic formula for q and σ .

Let the parameter “deplete” be the initial biomass as a fraction of k . What are the point estimates of r , k , and deplete. Show a plot of the observed and predicted index.

Draw the likelihood profile for r and then the likelihood profile for k . In both cases show the implied MSY for each value of r or k .

Repeat the calculations of part 2 just using the data from years 1 to 17. What advice would you give about the potential yield of the stock.

Using all the data calculate the likelihood profile on the stock size at the end of the last year of data.

Gulf of Georgia Herring catch and index of abundance

Year	Catch	Index of Abundance
1	41.44	105.9
2	47.12	122.5
3	7.76	226.7
4	63.04	152.9
5	62.32	173.5
6	62.96	108.4
7	62.08	72.6
8	24.80	114.9
9	48.88	140.9
10	69.92	145.8
11	34.96	120.3
12	67.44	87.1
13	70.40	129.8
14	81.60	136.6
15	63.20	77.9
16	44.96	72.2
17	43.04	109.2
18	6.24	76.1
19	0.64	135.6
20	0.96	227.6
21	2.00	225.8
22	11.90	207.1
23	10.80	155.5
24	7.00	203.2
25	8.39	215.3
26	15.30	195.1
27	20.20	196
28	27.40	164
29	22.60	206.3
30	8.89	211.4

1.6.1 Bootstrap

Open the file “bootstrap one.xls” and explore how it organized to provide 10 bootstrap estimates of the value of “rbar” with rbar broken into discrete levels of 10.

The key features of the top part of the sheet are the F columns with the residuals about the maximum likelihood fit and the H column which contains the original biomass surveys.

The G column contains the random numbers, drawn using the OFFSET command. For each value of “Random Index” in column F3, a different set of random numbers are put in the G column, and new values of the bootstrap sample data set are generated.

A two dimensional table function is then used to determine the likelihood for each value of rbar and random_index. This table is located at A41:L51.

Next for each cell in the likelihood table we determine if it was the highest likelihood for that row of random_index values. These are stored as 0 or 1 in cells A56:L66.

Finally we determine which rbar value has the highest likelihood for each random index value. I have done this by doing a SUMPRODUCT of the vector of rbar and the vector of 0 or 1.

Task 1: Use a similar approach to calculate the bootstrap estimates for the stock biomass in 1996.

Hint: use the table function to make a table that gives the biomass in 1996 for each level of rbar and each random_index.

Task 2: Extend the spreadsheet you just constructed to do 100 bootstrap replicates

Task 3: Construct a frequency distribution of the bootstrap estimates of B1996.

Note: This spreadsheet only approximates the bootstrap distributions because we use discrete levels of rbar rather than using something like SOLVER to find the maximum likelihood estimate of rbar for each bootstrap data set.

2 Bayesian estimation with a simple model and integration by direct search

2.1 Purpose

Perform the Bayesian parameter estimation in a simple one-parameter model: a population with constant recruitment and known survival. “States of nature” are recruitment levels, R . Calculate likelihood and posterior of R by direct search. Evaluate the effects of prior information and information in data on the posterior distribution. Compare results assuming that the variability in the data is due to either observation or process error.

2.2 Themes

- Bayesian estimation
- Information content in data
- Information content in prior distribution
- Fitting process versus observation error models

2.3 Data

Population size (in numbers) and catch (in numbers) for every year.

Year	Observed Catch	Biomass from surveys
1971	0	973
1972	6	1054
1973	6	581
1974	6	775
1975	18	942
1976	38	667
1977	68	681
1978	6	547
1979	19	554
1980	22	636
1981	27	657
1982	26	534
1983	32	671
1984	42	681
1985	36	723
1986	86	666
1987	162	619
1988	243	508
1989	198	461
1990	186	306
1991	179	267
1992	155	238
1993	133	361
1994	133	340
1995	105	457
1996	155	329

2.4 The model

Variables and parameters

N_t^o	Observed population in time t
N_t	Predicted population in time t
C_t	Observed harvest in time t
S	Finite survival rate (annual)
R	Average annual recruitment
σ_w^2	Variance of the process error
σ_v^2	Variance of the observation error
n	Number of observations

2.4.1 Dynamic model

$$N_{t+1} = sN_t + R - C_t \quad (1)$$

2.4.2 Initial Conditions

$$N_0 = \frac{R}{(1-S)} \quad (2)$$

2.4.3 Observation error

(variability due to mismeasurement of true state of nature)

$$N_t^o = N_t \exp(v_t); \quad v \sim N(0, \sigma_v^2) \quad (3)$$

2.4.4 Adding process error

(comprises all the variability in the state of the population produced by factors not included in the model).

$$N_{t+1} = [sN_t + R - C_t] \exp(w_{t+1}); \quad w \sim N(0, \sigma_w^2) \quad (4)$$

2.5 Estimation

2.5.1 Likelihood for observation error

The individual observation errors are

$$v_t = \ln \left(\frac{N_t^o}{N_t} \right)$$

The likelihood of the observation error is

$$P(v_t|R) = \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{v_t^2}{2\sigma_v^2}\right]$$

and the likelihood of all observation errors is therefore the product of the likelihoods for all years.

$$P(v|R) = \prod_t \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{v_t^2}{2\sigma_v^2}\right]$$

Finally we note that we since the v 's are the traditional sum of squares, we can convert the sum of squares to the likelihood as follows

$$SSQ = \sum_t v_t^2$$

$$P(v|R) \approx \exp\left(-\frac{SSQ}{2\sigma_v^2}\right)$$

The key to this simplification is that the first term of the likelihood is constant and therefore does not affect the likelihood. Assuming observation error, we can now refer to the likelihood with reference to the observed data.

$$P(N^0|R) \approx \exp\left(\frac{SSQ}{2\sigma_v^2}\right)$$

2.6 Prior

We assume that the prior is normally distributed

$$P(R) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left[-\frac{1}{2\sigma_R^2}(R - \bar{R})^2\right]$$

2.7 Posterior

$$P(R_x|N^0) = \left(\frac{P(N^0|R_x)P(R_x)}{\sum_{n_x} P(N^0|R_x)P(R_x)} \right)$$

The posterior distribution of R is given from Bayes law, where x is an interval in the recruitment space.

2.8 Laboratory

Build a spreadsheet to calculate the prior, the likelihood and the posterior for R , the average annual recruitment. Assume that:

- a) survival is known ($S=0.7$)
- b) only observation error is present ($\sigma_w^2 = 0$) and variability due to observation is known ($\sigma_v^2 = 0.16$).
- c) population in 1971 is at equilibrium ($R / (1 - S)$)
- d) prior is normal with $\bar{R} = 220$ and $\sigma_R^2 = 900$

Use the template “lab2-1 template” as a starting point and do things in the following steps

1. fill in the initial biomass in 1971 at location D13 using formula (2) for the initial conditions.
2. fill in the predicted biomass in 1972 in cells D14. To prevent the numbers from going negative if the R is set too low to support the observed catches, you should type in the formula as “=MAX(1,D13*s+rbar-B13)”
3. copy this formula down for all years.
4. type in the formula for the v_t^2 for 1971 in E13, and then copy down for all years.
5. enter the formula for the total sum of squares at B6
6. We are now going to use a 1 dimensional table to calculate the SSQ and then likelihood for a range of R values given in A43:A85. Begin this by putting =SSQ in cell b43. This is the column heading for a 1 dimensional table.
7. Select the range A43:B85, enter “Data Table” and choose rbar to change in the lower box. This will put in the B column the value of SSQ that is obtained for the values of rbar in the A column.
8. Convert the SSQ to the likelihoods in C44:C85. The variance of observation errors is represented by cv^2 in the template.
9. Put the normalized likelihood (taken by dividing by the sum of all likelihoods) in D44:D85.
10. Calculate the normal likelihood density function for the prior in E44:E85. The formula in E44 should be “=NORMDIST(A44,mean,standev,)”
11. Normalize the prior distribution by dividing by its sum and put this in F44:F85.
12. Calculate the raw (not normalized) posterior distribution in G44:G85 by multiplying the normalized likelihood times the normalized prior.
13. Normalize the posterior and put this in H44:H85.

14. Graphs are automatically generated and should show the fit to the data on one graph, and the prior, data and posterior on another.

2.9 Homework

Modify the spreadsheet built in the lab to fit the same model to the same data, but this time with a process error assumption (make $\sigma_w^2 = 0.16$ and $\sigma_v^2 = 0$). Keep all other assumptions unchanged and compare likelihood, prior, and posterior with those obtained in the lab.

3 Bayesian estimation with a two parameters and integration by direct search.

3.1 Purpose

Perform the Bayesian parameter estimation in a two-parameter model: a population with constant recruitment and known survival. Unlike laboratory 2, we now have only an index of abundance, and must estimate the scaling factor between the index of abundance and the real biomass. “States of nature” are recruitment levels, R and values of q . Calculate likelihood and posterior by direct search. Evaluate the effects of prior information and information in data on the posterior distribution.

3.2 Themes

Bayesian estimation with two variables
Information content in data
Information content in prior distribution

3.3 Data

Population size (index of abundance) for some years and catch (in numbers) every year.

3.4 The model

Variables and parameters

I_t^o Observed index of abundance in time t
 N_t Predicted population biomass in time t
 C_t Observed harvest in time t
 S Finite survival rate (annual)
 R Average annual recruitment
 σ_v^2 Variance of the observation error
 n Number of observations
 q the scaling factor between index of abundance and abundance
 I_t the index of abundance at time t

3.4.1 Dynamic model

$$N_t = S N_{t-1} + R - C_{t-1} \quad (1)$$

3.4.2 Initial Conditions

$$N_0 = \frac{R}{(1-S)} \quad (2)$$

3.4.3 Observations

$$I_t = qN_t \quad (3)$$

3.4.4 Observation error

(variability due to mismeasurement of true state of nature)

$$I_t = qN_t \exp(v_t); \quad v_t \approx N(0, \sigma_v^2) \quad (4)$$

3.5 Estimation

3.5.1 Likelihood for observation error

The individual observation errors are

$$v_t = \ln\left(\frac{I_t^o}{I_t}\right) \quad (5)$$

The likelihood of the observation error is

$$P(v_t|R) = \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{v_t^2}{2\sigma_v^2}\right] \quad (6)$$

and the likelihood of all observation errors is therefore the product of the likelihoods for all years.

$$P(v|R, q) = \prod_t \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{v_t^2}{2\sigma_v^2}\right] \quad (7)$$

Finally we note that we since the v's are the traditional sum of squares, we can convert the sum of squares to the likelihood as follows

$$SSQ = \sum_t v_t^2 \quad (8)$$
$$P(v|R, q) \approx \exp\left(-\frac{SSQ}{2\sigma_v^2}\right)$$

The key to this simplification is that the first term of the likelihood is constant and therefore does not affect the likelihood. Assuming observation error, we can now refer to the likelihood with reference to the observed data.

$$P(N^0|R, q) \approx \exp\left(\frac{SSQ}{2\sigma_v^2}\right) \quad (9)$$

3.5.2 Priors

We assume that the prior on R is normally distributed

$$P(R) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left[-\frac{1}{2\sigma_R^2}(R - \bar{R})^2\right] \quad (10)$$

We consider two possible priors for q, either it is uniformly distributed, or it is log uniform, that is

if uniform :
$$P(q_i) = \frac{1}{n_{\text{int}}} \quad (11)$$

if Log-uniform :
$$P(q_i) \approx \frac{\ln(q_{i+1}) - \ln(q_i)}{\sum_{n_{\text{int}}} \ln(q_{i+1}) - \ln(q_i)} \quad (12)$$

Further we assume that the priors of R and q are *independent*.

3.5.3 Posterior

Let us define H_k as hypothesis combining a specific R_i and q_j .

$$P(H_k | I) = \left(\frac{P(I|H_k)P(H_k)}{\sum [P(I|H_k)P(H_k)]} \right) \quad (13)$$

$$P(H_k) = P(R_i)P(q_j)$$

3.6 Laboratory

This spreadsheet is very similar to the one we built in laboratory 2, except that we are initially adding the complexity that q must be estimated. Again we will assume that:

- a) survival is known ($S=0.7$)
- b) only observation error is present ($\sigma_w^2 = 0$) and variability due to observation is known ($\sigma_v^2 = 0.16$).
- c) population in 1971 is at equilibrium (Eq. 2)

Use the template “twogrid part I template” as a starting point. This sheet already has the calculations to obtain the likelihood for a specific value of R and q completed. Your task is going to be to calculate the *joint likelihood of R and q* values using a TABLE function, then compute the *Bayes posterior joint and marginal distributions*.

Do the following steps.

1. Enter the formula for the total likelihood in cell B45.
2. Next build a 2 dimensional table in cells B45:U77, specifying that the row input cell is *qpar* and the column input cell is *rbar*. This assures that the q values shown in row 45 and the R values used in column B will be properly used, and the entries in this table will be the likelihood for each combination of R and q .
3. Now note that if you go to the sheet “Data” and change column E from a value of 1 to 0, then the simulated index of abundance for that year will be omitted. If you set column E equal to 1 for each year, you will give the program data every year and a highly informative data sequence. If you set most of the entries in column E to 0 then you provide less information. If you don’t give the estimation any data points after

the large catches begin, then you are providing very poor information to the estimation.

4. You have constructed a very large two dimensional table, that should take 5-15 seconds to recalculate. For this reason we have set the “Tools Options Calculations” option to manual recalculation for Tables, so that you must press F9 to recalculate the table. Otherwise there would be a many second delay every time you changed any cells. EXCEL will warn you with a “calculate” warning at the bottom of the spreadsheet when you have changed a value and need to press F9 to see the correct answer.
5. Now load the spreadsheet “twogrid part II template”. This spreadsheet looks very much like “twogrid part I template” but instead of having the big TABLE function to calculate the likelihoods, it has a “LINK” to the values in “twogrid part I template”. In this spreadsheet we are going to add the priors to produce the Bayesian posteriors.
6. You should still have “twogrid part I template” loaded. In your part II sheet, in cells A83:A114 add the prior distribution for R using equation 10, and note that the mean of the prior is at \$B\$10 and the standard deviation is at \$B\$11. You can use the EXCEL function NORMDIST to calculate the normal probability density if you wish.
7. Calculate the likelihood profile on R in cells V46:V77 by finding the maximum value of the likelihood for each row of the likelihood table. Then calculate the normalized likelihood by dividing by the SUM(V\$46:V\$77) and put this in W46:W77. This will be used later to graph the likelihood profile against the Bayes posterior.
8. At cell C81 insert the prior for q as follows “=IF(qprior=1,1,LN(D82/C82))”. This will put in a non-normalized uniform prior (1 for all cells) if qprior=1, and a non-normalized log uniform if qprior is not 1. Replicate this formula to the right as far as U81 for all columns.
9. Now at C83 we want the numerator of the Bayes posterior, that is the likelihood (from C46) times the prior for R and the prior for q. Now replicate this formula down and right to U114.
10. Next we want to normalize each element of the Bayes posterior. So at C118 divide the numerator of the Bayes posterior (C83) by the sum over all hypotheses (SUM(\$C\$83:\$u\$114)) and replicate this formula down and right to fill in the matrix.
11. Now calculate the marginal posterior distribution for q in cells c150:U150. This is simply the sum over all values of R for a particular value of q.
12. Finally, calculate the marginal posterior distribution for R in cells V118:v149, again by summing across the rows of the posterior for all values of q.
13. STOP and read carefully the next two items. We are dealing with **PRIORS!!** In C151:U151 calculate the **normalized prior** for q by dividing the prior for q by the

total of all priors for q . This is used for plotting –this is not the normalized posterior but refers back to the prior on q .

14. In W118:W149 calculate the **normalized prior** for R .
15. This spreadsheet should be working at this point and in the sheet called “posteriors” you should see a graph of the marginal posterior distribution for each parameter (vertical bars) and the prior for the parameter (thin line).
16. When you loaded the two worksheets the index data were available for all years. Now explore two different priors on R , first a non-informative one with mean 300 and standard deviation 100 (Part II – A). Look at the graphs in “posteriors” and see that the prior has almost no influence on the posterior for R . Next try a prior of 350 with a standard deviation of 50 (Part II – B). If for some reason you had a strong prior such as this, it would shift the posterior away from what the data are saying.
17. Now we want to explore what happens with a less informative data series. This will involve recalculation of the large likelihood table, and you will need to close the part II worksheet, so that it doesn’t recalculate for every step of the part I. Worksheet. So just say “file close” for part II.
18. Now go into the “data” sheet for part I and set the “to use” index for 1985 and all years after equal to 0, so that no indices are available after 1985. Then go back to the “main” sheet and press F9. This will cause a recalculation. This data series provides no information about what happened after the increase in catches.
19. Reopen part II and set the prior for totally uninformative mean=300 standard deviation 300, and the prior for q equal to uniform “qprior=1”. Note that while the likelihood is almost totally flat beyond 210 (the minimum R required to generate the known catches), the posterior on R declines considerably, suggesting a smaller R than the data. This is because a uniform prior on q is informative on R . Set qprior=1 to provide a log-uniform prior on q . Then look at the posterior on R —it is now the same as the likelihood profile (Part II – C).

4 Decision theory/decision table with Monte Carlo simulation.

4.1 Purpose

Project a population model and build a decision table. States of nature are levels of λ , *i.e.* how productive the population is. Management options or controls are harvest levels (harvest rates or quotas). The outcomes to be explored are the average annual harvest, the variance in harvest and the probability of collapse. A deterministic model will be constructed first and a stochastic model will be built later for the analysis of persistence.

4.2 Themes

Basics of decision analysis
Average harvest, variability in harvest, and population persistence
Fixed quotas vs. fixed harvest rates
Effect of model structure on conclusions

4.3 The model

Variables and parameters

N_t Population size in time t
 C_t Number harvested in time t
 a, b Parameters of Beverton-Holt model
 λ Finite rate of population growth
 K Population carrying capacity
 h Annual harvest rate

4.3.1 Annual quota

Beverton-Holt model

$$N_{t+1} = \frac{a N_t}{1 + b N_t} \quad (1)$$

Reparameterization in terms of annual growth (λ) and carrying capacity (K)

$$N_{t+1} = \frac{\lambda K N_t}{K + (\lambda - 1) N_t} \quad (2)$$

Extra credit:

Show that (2) is a reparameterization of (1).

4.3.2 Adding harvest

$$N_{t+1} = \frac{\lambda K (N_t - C_t)}{K + (\lambda - 1)(N_t - C_t)} \begin{cases} C_t = h N_t & \text{if constant harvest rate} \\ C_t = \text{Min} (Q, N_t) & \text{if constant quota} \end{cases}$$

4.3.3 Adding stochasticity

$$N_{t+1} = \frac{\lambda K (N_t - C_t)}{K + (\lambda - 1)(N_t - C_t)} e^{\mathcal{E}} ; \quad \mathcal{E} \sim N(0, \sigma^2)$$

4.3.4 Outputs

$$\text{Average harvest} = \text{AVG}(C_{191}:C_{200})$$

$$\text{Variability in the harvest} = \frac{SD(C_{191}:C_{200})}{\text{AVG}(C_{191}:C_{200})}$$

$$\text{Average population size} = \text{AVG}(N_{191}:N_{200})$$

$$\text{Prob. below } 1/10 \text{ K} = \frac{\#\{N_t < 0.1K\}}{\#\text{years}}$$

$$W_{-avg_p} = \sum_{i=1}^n O_{\lambda_i,p} P_{\lambda_i}$$

where, W_{-avg_p} is the weighted average of the desired output (O) under exploitation level p , and P_{λ_i} is the probability associated the state of nature λ_i .

4.4 Laboratory

Part I. Build a spreadsheet to project the model in the future and evaluate the equilibrium harvest (in year 200) when the population is harvested **a**) by setting a constant quota (between 0 and 880 individuals per year) or **b**) a constant harvest rate (between 0 and 27.5% per year). Build a decision analysis table by assuming that $K = 10,000$. The different “states of Nature” are represented by different values of λ , which can take any value between 1.05 and 1.35 (between 5 and 35% maximum annual growth) with the same probability.

- Use the file “Temp_Lab1.xls” to do the projections.
- Complete the columns of NUMBERS and HARVEST by using the above given equations.
- Do a two dimensional table on the equilibrium harvest for each harvest policy.

Part II. Add stochasticity to the model by assuming that $\sigma = 0.15$. Build a decision analysis table for *average harvest*, *variability in harvest* and *probability of the population dropping below 1,000 individuals*.

- Generate a column of 200 random number ($\sim N(0, \sigma^2)$) by using the excel command from Tools/Data Analysis/ Random Number Generation/. Use the seed “111”, so that you get results comparable to the answer sheet.
- Multiply the abundance vector by e^ε .
- Generate three tables for each harvest policy.

4.5 Homework

Modify the spreadsheet built in II to represent the dynamics of the population with a Ricker model:

$$N_{t+1} = \lambda(N_t - C_t) \exp\left[-\frac{\ln(\lambda)}{K}(N_t - C_t)\right]$$

Repeat the analysis in II and compare the results in terms of average harvest, variability in harvest and population persistence. Use same productivity and same carrying capacity.

5 Integration using MCMC and SIR

5.1 Purpose

Introduces different integration methods and demonstrates the process and fundamental properties of these algorithms. Link integration methods with decision analyses formally by presenting an example application to Westcoast POP.

5.2 Themes

Integration methods
High dimensional problems

5.3 Model

Statistical catch-age model
Includes stock-recruitment curve to assess stock productivity
Projection/policy analyses natural extension

5.3.1 Files

5.4 Lab

Begin with “SIR no macros.xls” showing basic SIR steps. Key elements

1. drawing from the prior
2. evaluating the likelihood at that point
3. integrating across different parameter values by summing using pivot table

Now, what does it take to look at the posterior of B2020? This is shown in “SIR with projections”

But lets modify the old sheet

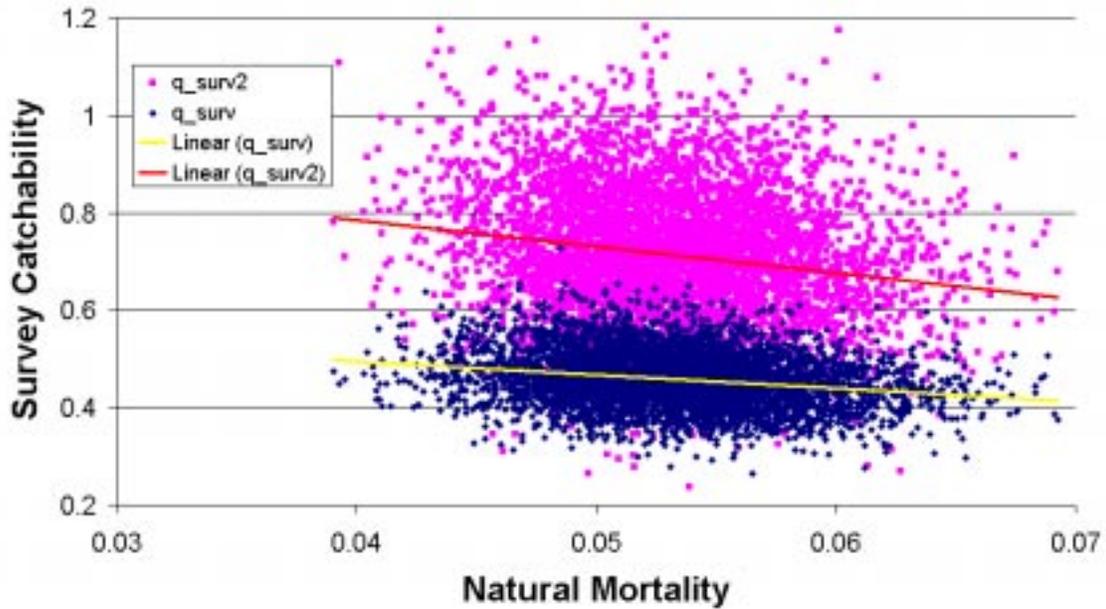
1. Project ahead to 2020 - need to add harvest rate
2. Add output of B2020 to table, also add steps 1-100 for pivot table

Next lets look at “MCMC with projections” for a simple example of MCMC with no macros.

5.5 Dealing high-dimensional problems: westcoast POP example

So far, the integration and decision tables have dealt with simplistic models. As one can imagine, presenting higher dimensional problems into a simple, summary form useful for fishery management consequences can be problematic. For example, the 1998 Pacific Council POP assessment included over 450 parameters (including projection variables and parameters of interest). Distilling this down to a simple “productivity” parameter to show alternative “states of nature” was impossible (without making the usual large number of assumptions). In this case, MCMC was used to obtain a posterior distribution (using AD Model Builder) to develop a reasonable representation of the posterior distribution. This was done by simply extracting a sub-sample of the original 1 million parameter draws to get a final sample of 7,500 parameters. Using these 7,500 parameter sets, we selected 39 different parameters of interest for generating summaries. For

example, the following figure shows the joint distribution of two different survey catchability coefficients (vertical axis) and natural mortality rate (horizontal axis).



Since the key management issue was how different actions (harvest levels) in the future may affect stock rebuilding, we simply sorted the posterior distribution and categorized each parameter set as falling within one of three different levels of current stock size (high, medium, and low). These are shown in the table below. Also shown are marginal distributions of future stock size under different harvest level scenarios relative to some benchmark stock sizes (the 1998 level and the target).

Decision table showing outcomes of alternative harvest rate applications (rows) by different 1998 stock size hypotheses (columns). The labels low, mod, and high represent the lower, middle, and upper third quantile of the 1998 stock size. For these columns, the values shown represent the expected outcome within that quantile.

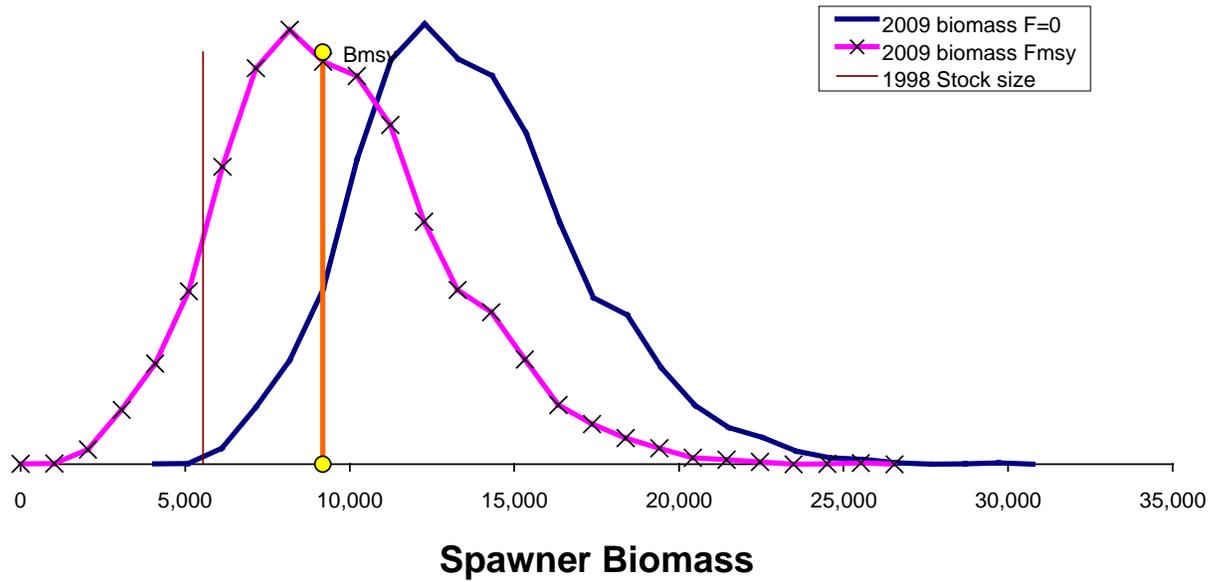
1998 Stock Size			
Low	Mod	High	Expected Value
3,970	5,384	7,275	5,543

Policy	1999 Harvest (tons)
F=0	0
Fmsy	794
F40	695
F35	834
F30	1,007

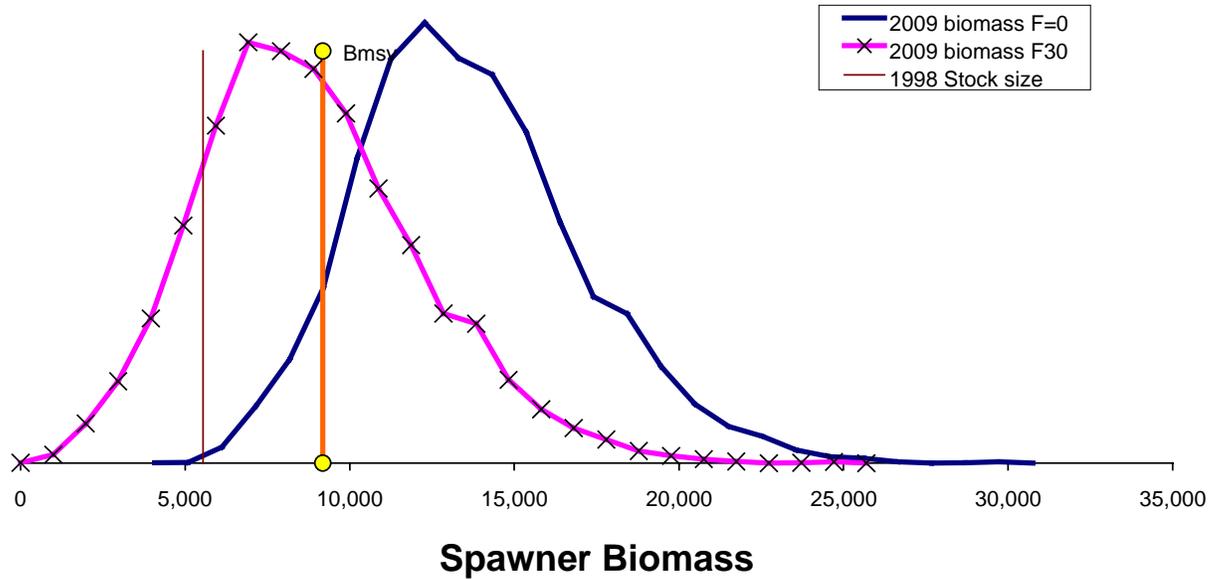
2009 Stock Size				
Policy	Low	Mod	High	Expected Value
F=0	10,160	12,966	16,295	13,140
Fmsy	6,097	8,940	12,299	9,112
F40	6,500	9,338	12,694	9,511
F35	5,942	8,786	12,146	8,958
F30	5,307	8,158	11,523	8,329

Ratio 2009 / 1998 Stock Size				
Policy	Low	Mod	High	Expected Value
F=0	256%	241%	224%	237%
Fmsy	154%	166%	169%	164%
F40	164%	173%	174%	172%
F35	150%	163%	167%	162%
F30	134%	152%	158%	150%

Ratio 2009 / Bmsy				
Policy	Low	Mod	High	Expected Value
F=0	105%	142%	186%	143%
Fmsy	63%	98%	140%	99%
F40	67%	102%	145%	104%
F35	62%	96%	138%	98%
F30	55%	89%	131%	91%



Probability distributions of projected female spawner biomass in the year 2009 under F_{msy} harvest compared to F=0 harvests. Vertical lines are reference points.



Probability distributions of projected female spawner biomass in the year 2009 under $F_{30\%}$ harvest compared to F=0 harvests. Vertical lines are reference points.

5.6 Homework

Identify as many pitfalls in the type of decision analysis used for WC POP as you can think of.

6 References

- Arnold, S.F. 1990. Mathematical statistics. Prentice Hall Inc. Englewood Cliffs, New Jersey. p. 636.
- Berger, J.O. 1985. Statistical decision theory and Bayesian analysis. (2nd Ed). Springer-Verlag, New York, NY, 617 p.
- Box, G.E.P., and Tiao, G.C. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, Mass. 588p.
- Fournier, D.A. and C.P. Archibald. 1982. A general theory for analyzing catch-at-age data. *Can. J. Fish. Aquat. Sci.* 39:1195-1207.
- Francis, R.I.C.C. 1992. Use of risk analysis to assess fishery management strategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. *Can. J. Fish. Aquat. Sci.* 49: 922-930.
- Gelfand, A.E. and Smith, A.F.M. 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85: 398-409.
- Gelman, A, J.B. Carlin, H.S. Stern, and D.B. Rubin. 1995. Bayesian data analysis. Chapman and Hall. London. 526 p.
- Geweke, J.F. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57: 1317-1340.
- Givens, G.H. 1993a. Contribution to discussion on the meeting on 'The Gibbs sampler and other Markov chain Monte Carlo methods'. *J. R. Statist. Soc. B* 55: 75-76.
- Givens, G.H. 1993b. A Bayesian framework and importance sampling methods for synthesizing multiple sources of evidence and uncertainties linked by a complex mechanistic model. Ph.D. thesis, University of Washington, Dept. of Statistics, Seattle, WA.
- Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* 57:97-109.
- Hilborn, R. and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman and Hall, New York, N.Y. 570 p.
- Hilborn, R. and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman and Hall, New York, N.Y. 570 p.
- Ianelli, J. and Heifetz, J. 1995. Decision analysis of alternative harvest policies for the Gulf of Alaska Pacific ocean perch fishery. *Fisheries Research* 24:35-63.

- Ianelli, J.N. and D.A. Fournier. 1998. Alternative age-structured analyses of the NRC simulated stock assessment data. *In* Restrepo, V.R. [ed.]. Analyses of simulated data sets in support of the NRC study on stock assessment methods. NOAA Tech. Memo. NMFS-F/SPO-30. 96 p.
- Kinas, P.G. 1993. Bayesian statistics for fishery stock assessment and management. Ph.D. thesis, University of British Columbia, Dept. of Statistics, Vancouver, B.C.
- Kinas, P.G. 1996. Bayesian fishery stock assessment and decision making using adaptive importance sampling. *Can. J. Fish. Aquat. Sci.* 53: 414-423.
- McAllister, M.K. 1995. Using decision analysis to choose a design for surveying fisheries resources. Ph.D. thesis, University of Washington, School of Fisheries, Seattle, WA.
- McAllister, M.K., and Pikitch, E.K. 1996. A Bayesian approach to choosing a design for surveying fishery resources: application to the eastern Bering Sea trawl survey. *Can. J. Fish. Aquat. Sci.* 53:.
- McAllister, M.K., and Ianelli, J.N. 1996. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. *Can. J. Fish. Aquat. Sci.* 53:.
- McAllister, M.K., Pikitch, E.K., Punt, A.E., and Hilborn, R. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Can. J. Fish. Aquat. Sci.* 51: 2673-2687.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087-1092.
- Myers, R.A., Bridson, J., and Barrowman, N.J. 1993. Summary of worldwide stock and recruitment data. *Can. Tech. Rep. Fish. Aquat. Sci.* 2024:327p.
- Nash, J.C. 1990. Compact numerical methods for computers: linear algebra and function minimization. 2nd Edn. Adam Higler Ltd. New York, NY. 278p.
- Newton, M.A. and Raftery, A.E. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J.R. Statist. Soc. B.* 56: 3-48.
- Oh, M.S. and Berger, J.O. 1992. Adaptive importance sampling in Monte Carlo integration. *J. Statist. Comput. Simul.* 41: 143-168.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. 1992. Numerical Recipes in C. Second Ed. Cambridge University Press. 994 p.
- Punt, A.E., Butterworth, D.S., and Penny, A.J. 1993. Stock assessment and risk analysis for the South Atlantic population of albacore (*Thunnus alalunga*). ICCAT Doc. SCRS/93/83.

- Raftery, A.E. and Lewis, S.M. 1992. How many iterations in the Gibbs sampler? In Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, April 15-20, 1991. Edited by J.M. Bernardo, J.O. Berger, P.A. David, and A.M. Smith. Clarendon Press, Oxford. pp.763-773.
- Raftery, A.E., Givens, G.H., and Zeh, J.E. 1995. Inference from a deterministic population dynamics model for Bowhead whales. *J. Am. Stat. Assoc.* 90: 402-416.
- Ripley, B.D. 1987. Stochastic simulation. John Wiley and Sons Inc. New York, NY. 237 p.
- Rubin, D.B. 1987. Comment on 'The calculation of posterior distributions by data augmentation'. *J. Amer. Statist. Ass.* 82:543-6.
- Rubin, D.B. 1988. Using the SIR algorithm to simulate posterior distributions. In Bayesian Statistics 3: Proceedings of the Third Valencia International Meeting, June 1-5, 1987. Edited by J.M. Bernardo, M.H. Degroot, D.V. Lindley, and A.M. Smith. Clarendon Press, Oxford. pp. 385-402.
- Smith, A.F.M. 1991. Bayesian computational methods. *Phil. Trans. R. Soc. Lond. A.* 337: 369-386.
- Thompson, G.G. 1992. A Bayesian approach to management advice when stock-recruitment parameters are uncertain. *Fish. Bull.* 90:561-573.
- van Dijk, H.K. and Kloek, T. 1983. Monte Carlo analysis of skew posterior distributions: an illustrative economic example. *The Statistician* 32: 216-233.
- Walters, C.J. and Ludwig, D. 1994. Calculation of Bayes posterior probability distributions for key population parameters: a simplified approach. *Can. J. Fish. Aquat. Sci.* 51: 713-722.
- Walters, C.J. and Punt, A.E. 1994. Placing odds on sustainable catch using virtual population analysis and survey data. *Can. J. Fish. Aquat. Sci.* 51: 946-958.
- West, M. 1992. Modeling with mixtures. In Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, April 15-20, 1991. Edited by J.M. Bernardo, J.O. Berger, P.A. David, and A.M. Smith. Clarendon Press, Oxford. pp. 503-524.
- West, M. 1993. Approximating posterior distributions with mixtures. *J.R. Statist. Soc. B.* 55: 409-422.
- Wilderbuer, T.K., Walters, G.E., and Bakala, R.G. 1992. Yellowfin sole, Pleuronectes asper, of the eastern Bering Sea: biological characteristics, history of exploitation, and management. *Marine Fisheries Review* 54: 1-18.

Some Windows / Excel tips

Navigation

Global

ALT Tab	Switch between applications
CTL F6	Switch between sheets within an application
ALT SPACE	Access minimize/maximize/restore global (application) menu (underlined letters map to functionality)
ALT -	Access minimize/maximize/restore local (child window) menu (underlined letters map to functionality)
ALT F4	Quit current application

In Excel

CTL Home	Go to upper corner
CTL End	Go to lower corner
CTL left arrow	Go to leftmost block of current of data
CTL right arrow	Go to rightmost block of current of data
CTL up arrow	Go to uppermost block of current of data
CTL down arrow	Go to lowermost block of current of data
CTL page down	Go to next sheet (w/in current file)
CTL page up	Go to previous sheet
SHIFT CTL Home	Select to upper corner
SHIFT CTL End	Select to lower corner
SHIFT CTL left arrow	Select to leftmost block of current of data
SHIFT CTL right arrow	Select to rightmost block of current of data
SHIFT CTL up arrow	Select to uppermost block of current of data
SHIFT CTL down arrow	Select to lowermost block of current of data
ALT mouse object	Snaps object (e.g., chart) to align with cells
ALT Insert Name Create	Name cells by left column or top row values

Cut n paste:

CTL c (CTL INSERT)	Copy current selection to clipboard
CTL x (SHIFT DELETE)	Cut current selection to clipboard
CTL v (SHIFT INSERT)	Paste clipboard contents (works for many things, e.g., charts etc)
CTL SHIFT Enter	Output results of arrays (e.g., multi-cell values)
CTL mouse click/drag object	Lifts copy of object

Some common functions:

Sum(arr1)	Sums contents of array (or named array) arr1
Sumproduct(arr1,arr2)	Sums the element-wise product of two vectors (or matrices) arr1 and arr2
Sumxmy2(arr1,arr2)	Returns the sums of squared differences between two vectors (or matrices) arr1 and arr2 (norm squared of differences)
Sumxmy2(ln(arr1),ln(arr2))	Returns the sums of squared differences between logarithms of two vectors (or matrices) arr1 and arr2 (norm squared of differences)
Rand()	Uniform random number (0,1)
NORMINV(prob,mean,std)	Inverse of normal distribution. Prob is probability level), std is standard deviation.
NORMINV(RAND(),0,1)	Returns a normal 0,1 random variable